



EIXO 1 – AUDITORIA PÚBLICA CENTRADA NO CIDADÃO

A KNOWLEDGE GRAPH FOR MONITORING AND CONTROLLING PUBLIC POLICIES

Álvaro Augusto Bastos de Carvalho

Formado em Processamento de Dados pela Universidade de Brasília (UNB-1986) e Bacharel em Direito pela Universidade Federal do Ceará (UFC-1996). Possui diversos cursos de especialização em auditoria e informática. Ingressou no Tribunal de Contas da União (TCU) através de concurso público em 1987 e já exerceu as funções de Assessor, Diretor e secretário/Gestor da Secretaria de Controle Externo do Tribunal de Contas da União no Ceará (Secex/CE). Participou de várias auditorias como coordenador e membro de equipe, como também já instruiu processos relevantes de todos os tipos. Atualmente é Auditor Federal de Controle Externo do Tribunal de Contas da União. Tem muita experiência nas áreas de Ciências da Computação e de Auditoria. Em 12/2024, concluiu o Mestrado em Informática Aplicada, da Universidade de Fortaleza. Atualmente, cursando o Doutorado em Informática Aplicada, na Universidade de Fortaleza.

Ramir Holanda Filho

Pós-Doutor em Computação pela Sorbonne Université-Pierre et Marie Curie (França - 2020). Doutor em Ciência da Computação pela Universitat Politècnica de Catalunya (Espanha - 2005). Atualmente é professor titular da Universidade de Fortaleza - UNIFOR na qual é membro permanente dos programas de pós-graduação em Informática Aplicada (mestrado e doutorado) e do mestrado profissional em administração. Possui mais de 130 trabalhos publicados em conferências e periódicos nacionais e internacionais. Tem experiência na área de Ciência da Computação, com ênfase em Ciência de Dados, atuando principalmente nos seguintes temas: Redes Complexas, Grafos de Conhecimento, Graph Neural Networks, Inteligência Artificial, Segurança.

Os autores signatários encaminharam um *paper* (artigo) intitulado *A Knowledge Graph for Monitoring and Controlling Public Policies* (Grafo de Conhecimento para monitorar e controlar políticas públicas) para a 13ª Edição do *International Conference on Complex Networks and Their Applications*, o qual foi aceito e apresentado em 14/12/2024, em Istambul, Turquia, tendo sido publicado como capítulo de um livro da Série *Studies in Computational Intelligence*, da Editora Springer Nature¹.

Objetivando contribuir para melhorar as ações de fiscalização das políticas públicas, o primeiro autor apresentou o projeto, em 11/2/2025, ao E.TCU, em particular à SETID e à SEINC, que aquiesceram às propostas e patrocinaram o desenvolvimento do projeto.

Descreve-se, a seguir, um resumo do projeto e o andamento de sua implementação no TCU, pois o tema guarda correlação com diversos eixos do presente Seminário, vez que propõe a criação de

¹ (https://link.springer.com/chapter/10.1007/978-3-031-82435-7_10; volume III, p. 115-126)



um grafo de conhecimento que tem o cidadão como peça central para o controle das políticas públicas, com monitoramento de suas principais interações com o Estado brasileiro. O projeto está sendo desenvolvido sob a supervisão da Seinc-Adjunta.

INTUIÇÃO SOBRE O ARTIGO

Sabe-se que as Políticas Públicas são dirigidas aos cidadãos de seus respectivos países. Contrapondo a visão tradicional, que normalmente implementa a análise de políticas públicas no sentido do Estado para o cidadão, o presente projeto buscou verificar a viabilidade de criação de um sistema de informação com um viés reverso: do cidadão para o Estado, enfatizando como o cidadão está sendo atendido pelo Estado em todos os seus aspectos. Se tal concepção for viável, o projeto busca responder quais são os requisitos necessários, como deve ser a arquitetura do banco de dados e se haveria benefícios nessa abordagem, como, por exemplo, a possibilidade de detectar fraudes na concessão de benefícios sociais. Preponderantemente, os sistemas de controle dessas políticas públicas são desenvolvidos em um banco de dados de modelo relacional (SQL). As tabelas relacionais têm potencial de formar consultas poderosas por meio de junções entre elas, que, se forem de volume significativo de dados, podem consumir muitos recursos computacionais e geralmente não estão disponíveis *online*. Essa modelagem não permite uma visão completa da interação entre o Estado e o cidadão. Por várias razões técnicas, analisadas no projeto, o sistema de informação precisa ser desenvolvido em um banco de dados NoSQL baseado em grafos. O presente trabalho tem como objetivo verificar se um banco de dados em grafos permite uma modelagem tendo o cidadão como centro das políticas públicas. Por meio de dados públicos de três programas sociais de transferência de renda no Brasil, foi criado um grafo com 433 milhões de nós e arestas. O estudo permitiu concluir que a modelagem de políticas públicas x cidadãos é viável por meio de grafo de conhecimento, possui inúmeras vantagens em relação ao modelo tradicional de tabelas relacionais e apresenta a possibilidade de detecção de fraudes.

A Figura 1, ilustra, em linhas gerais, o *schema* do banco de dados baseado em grafos. Representa as conexões do cidadão com diversos sistemas da Administração pública federal (APF), como: agências que controlam barcos, aviões, automóveis; sistemas, como de óbitos, eleitoral, SUS, de



concessões de benefícios, bolsas de estudo; empresas e quadros societários; compras públicas; cadastros de informações sociais e financiamentos públicos.

Figura 1 – Schema do Knowledge graph



No modelo relacional, a perspectiva que se tem de um cidadão em particular é aquela do respectivo banco de dados, mas, para a obtenção de sua visão holística, é necessário desenvolver um sistema para coletar os dados dos inúmeros bancos de dados da APF e, para captação da visão integrada de vários sistemas, com inúmeras interconexões entre eles, a tarefa é ainda mais complexa. Já no banco de dados grafo, após identificado o cidadão, é instantânea a verificação de todas as suas conexões, permitindo uma visão abrangente do cidadão e dos inúmeros aspectos das políticas públicas.

Assim, o protótipo de *Knowledge Graph* (KG) da Figura 1 permite, por exemplo, que se façam auditorias de políticas públicas transversais, tais como, se existem cidadãos que:

1. morreram e estão recebendo benefícios sociais, ou, recebem benefícios sociais e são proprietários de carros, aviões ou barcos de valores significativos, incompatíveis com o benefício;
2. recebem benefícios sociais e são proprietários de empresas ou são empregados de outras empresas na RAIS (Relação Anual de Informações Sociais) com salários elevados, ou têm financiamentos vultosos de bancos públicos;



3. com salários baixos na RAIS, mas que são sócios de empresas com faturamentos elevados ou que contraíram financiamentos públicos significativos, o que pode ser indicativo de que são “sócios-laranjas”;

Conclui-se que os grafos são a alternativa adequada quando as complexas conexões entre os dados são tão importantes quanto os dados em si. Dessa forma, se houver uma aresta conectando os nós do grafo, a informação pretendida pode ser obtida, pois o KG é capaz de modelar, de forma simples, os objetos/entidades do mundo real e suas conexões.

Um caso prático da utilização de KG para detectar fraudes em aquisições públicas foi desenvolvido pelos signatários, apresentado no *The 15th International Conference on Ambient Systems, Networks and Technologies Networks (ANT)*, realizado em abr/2024, Hasselt University, Belgium (*Detecting Fraud in Public Acquisition of Brazilian Government with an Analytical Approach*)².

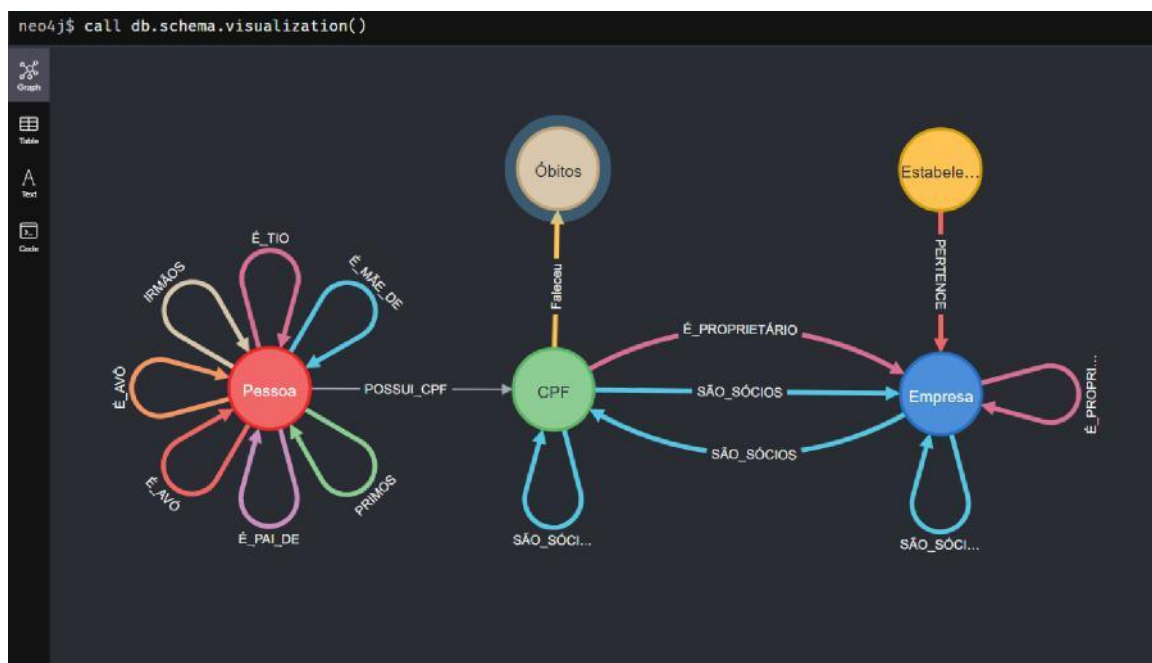
Posição atual do projeto

Como o projeto centra-se no cidadão, a primeira etapa foi identificar os cidadãos nas bases de dados federais: Cadastro da Receita Federal (274 milhões de CPFs), Cadastro Eleitoral (171 milhões de inscrições), Carteira Nacional de Habilitação (CNH; 84 milhões). Assim, foram identificados os cidadãos e relações de filiação (pais/mães x filhos) e inferidas as demais conexões familiares. Além disso, foram inseridas no grafo as empresas existentes no país (60 milhões) e suas conexões societárias.

O processamento desses *datasets* requer ferramentas adequadas para *Big Data*: o banco de dados baseado em grafos utilizado é o *Neo4J* (versão *Community*) e o pré-processamento de dados em *PySpark*. O grafo continua evoluindo e já conta com dois bilhões e duzentos milhões de nós e arestas. A Figura 2 exibe o *schema* atual do grafo.

Figura 2 – Schema atual

² <https://www.sciencedirect.com/science/article/pii/S1877050924012584>



CONCLUSÃO

A conexão das informações dos diversos bancos de dados do LabContas em KG permite uma visão holística do cidadão e o monitoramento contínuo de políticas públicas. A integração dos dados também permite o uso de inteligência artificial, tal como as GNN - *Graph Neural Networks* (redes neurais baseadas em grafos) para identificar irregularidades na gestão pública.

A base de dados consolidada no grafo, combinada com a utilização de Agentes de IA utilizando RAG - *Retrieval Augmented Generation* (Geração Aumentada por Recuperação) e o recente protocolo MCP – *Model Context Protocol* (Protocolo de Contexto de Modelo) poderão representar significativo avanço na automação de auditorias de políticas públicas transversais.

Conclui-se que o projeto dos grafos, tal como descrito, tem conexão com diversos eixos temáticos do Seminário: auditoria pública centrada no cidadão; competências e habilidades do auditor do futuro, com uma visão ampla das políticas integradas e uso de novas tecnologias de processamento de dados; abordagem inovadora na auditoria, com uso de

> Seminário Internacional

O FUTURO DA AUDITORIA PÚBLICA

DADOS, INOVAÇÃO E CIDADANIA

21 e 22 de outubro de 2025

Instituto Serzedello Corrêa (ISC), Brasília (DF)



banco de dados baseados em grafos; inteligência artificial e transformação digital na auditoria pública.